Structure *14*, 973–982, June 2006 DOI 10.1016/j.str.2006.04.003

Phasing RNA Polymerase II Using Intrinsically Bound Zn Atoms: An Updated Structural Model

Technical Advance

Peter A. Meyer,[1] Ping Ye,[1] Mincheng Zhang,[1] Man-Hee Suh,[1] and Jianhua Fu[1,*]
[1] Department of Molecular Biology and Genetics
Cornell University
221 Biotechnology Building
Ithaca, New York 14853

Summary

Macromolecular assemblies as large as RNA polymerase II (Pol II) can be phased by a few intrinsically bound Zn atoms, by using MAD experiments as described here. A phasing effectiveness of 570 aa/Zn is attained for Pol II. The resulting experimental, unbiased electron density map is of such quality that it confirms the existing crystallographic model and further reveals structural regions not shown by model phases, thus updating the Pol II model at three sites. The mechanistically important fork loop-1 element is observed to be ordered in the absence of nucleic acids, suggesting additional insights into the mechanisms that maintain the stability of the transcription ternary complex and allow its release. Furthermore, a computational experiment with simulated MAD data sets demonstrates that 1 Zn site is able to provide adequate experimental phase information for as many as 1100 amino acids of polypeptide, under the conditions of the current synchrotron and detector technologies.

Introduction

Large macromolecular assemblies carry out the functions of vital cellular processes, including transcription and translation. A hallmark of these large systems is the structural organization of many auxiliary factors around a tightly associated core apparatus whose molecular weight challenges the methods of structure determination. In the instance of transcription of protein-coding genes, the core enzyme of RNA polymerase II (Pol II) consists of 12 polypeptide subunits that assemble into a stable and defined structure greater than 500 kDa (Darst et al., 1991; Woychik and Young, 1990). To elucidate the three-dimensional (3D) structure of Pol II and gain insights into the core mechanism of pre-mRNA synthesis, a number of steps were taken, including an innovative crystallographic treatment. First, an efficient scheme was developed that permitted purification of large amounts of the polymerase directly from a beneficial yeast strain (Edwards et al., 1990, 1991; Thompson et al., 1990). A form of crystal that diffracted X-rays to 3.1 Å and yielded reliable reflection amplitudes to 3.5 Å was then obtained (Fu et al., 1999) after initial crystallization of the yeast polymerase (A.M. Edwards and R.D. Kornberg, personal communication). Next, acquisition of initial reflection phase values became critical but was hampered by the fact that individual heavy atoms could not be resolved in difference Patterson calculations (a necessary step of Multiple Isomorphous Replacement [MIR]; Crick and Maydoff, 1956) due to the overwhelming contributions from the massive protein. This second bottleneck (crystallization being the first) was removed by the use of heavy atom cluster molecules, particularly, a tungstate with 18 tungsten atoms and a tantalum bromide with 6 tantalum atoms (Fu et al., 1999). Direct determination of the crystallographic coordinates of the clusters led to the calculation of initial phases for the Pol II crystal (Fu et al., 1999). The significance of this technical breakthrough was 2-fold: first, a 3D electron density map of Pol II, which revealed the mobility of a structural domain with profound mechanistic implications (Cramer et al., 2001; Fu et al., 1999; Gnatt, 2002; Gnatt et al., 2001), was obtained (at 6 Å); and second, the initial phases allowed for the identification of useful derivative crystals of discrete heavy atoms—particularly, a family of iridium compounds represented by pentamethylcyclopentadienyliridium cloride was identified (Fu et al., 1999) and later provided major phasing power in the establishment of the atomic model (Cramer et al., 2000)—and thus defined a route to the final solution. A similar phasing strategy led to the crystal structure determination of the large subunit of bacterial ribosome (Ban et al., 1998, 1999).

Additional structural results obtained after the crystallographic achievement of the core Pol II (Cramer et al., 2001) relied exclusively on molecular replacement (MR) (Lattman, 1985; Rossmann, 2001) and model-derived phases for electron density calculations (Armache et al., 2003; Bushnell et al., 2002, 2004; Bushnell and Kornberg, 2003; Gnatt et al., 2001; Kettenberger et al., 2003, 2004; Westover et al., 2004a, 2004b). As dictated by the properties of crystallographic structure factors, model-based phases have limited effectiveness in displaying protein densities not present in the search model. For the Pol II system, this limitation of phasing power worsens for cocrystals between the polymerase and associated factor(s), particularly for factors that represent greater than 20% of the total mass of Pol II (P.A.M. and J.F., unpublished data). As novel cocrystals of subassemblies are being produced for the Pol II machinery (which consists of 5–6 protein factors in its basal apparatus and several tens of factors in its various regulated complexes; Hartzog, 2003; Kornberg, 1999; Roeder, 1996; Shilatifard, 2004), this phasing problem will become a general rate-limiting step of crystallographic investigation. Therefore, strategies for generating experimental phases will be critical for solving various complex structures of Pol II and its regulatory factors. In principle, ab initio phases can be produced by derivatizing a Pol II-factor(s) cocrystal with the heavy atom reagents employed previously, such as the clusters as well as the iridium and rhenium compounds (Cramer et al., 2000; Fu et al., 1999), but this process is generally time consuming.

We present a convenient approach here that exploits anomalous scattering signals from the Zn atoms that are intrinsically bound in the Zn binding motifs of Pol II.

*Correspondence: jf227@cornell.edu

Table 1. Statistics of MAD Diffraction Data from Native Crystals of Pol II

Crystal	Data Type	λ (Å)	E (KeV)	I/σ	Comp (%)	Multiplicity	d-Spacing	R_{sym}
A4X4	inflection	1.2825	9.6670	5.5	99.8	5.9	7.6 Å	0.117
	remote	1.2795	9.6900	5.8	99.9	6.9	6.9 Å	0.109
A3X6[a]	inflection	1.2825	9.6670	4.1	99.8	13.2	7.2 Å	0.154
	remote	1.2651	9.8000	4.3	99.8	12.4	7.3 Å	0.148
A3X7[a]	peak	1.2821	9.6700	4.7	85.5	9.3	4.1 Å	0.143
	peak	1.2821	9.6700	7.5	99.6	13.7	6.5 Å	0.084
	inflection	1.2834	9.6600	6.7	99.2	8.8	7.3 Å	0.104
	remote	1.2755	9.7200	7.5	99.2	8.8	7.2 Å	0.093
A2X10[a]	peak	1.2823	9.6680	7.3	100.	11.5	6.2 Å	0.090
A2X11	inflection	1.2834	9.6600	6.9	82.2	4.6	6.6 Å	0.088
	remote	1.2755	9.7200	6.8	82.6	4.7	6.5 Å	0.092
Master set (isotropic)					100.	67.5	6.5 Å	0.153
Master set (anisotropic)					85.2	26.8	4.1 Å	0.165

Data were collected at beamline X25 of NSLS.
[a] Included in the final Zn-MAD phase calculation (below).

Although there are only eight Zn atoms in total, their collective contribution to the Multiwavelength Anomalous Diffraction (MAD) (Hendrickson, 1991; Hendrickson and Ogata, 1997) effect could be measured reliably by using synchrotron radiation. To recover the relatively weak Zn anomalous signal atop the large protein amplitudes from the massive polymerase (>4500 residues), it was necessary to adopt a data collection strategy that utilized multiple crystals and hence multiple phase sets for removing phase ambiguity. Such a treatment resulted in an experimental map that revealed definitive electron densities for three regions of Pol II that were disordered in the previous X-ray structures, although one such region was previously observed to be localized in the presence of nucleic acids, and thus allowed an update of the model. The Zn-MAD approach is applicable to structure determinations of any Zn-containing macromolecular assemblies whose partial structure is available, and it will aid future crystallographic efforts on complex structures of Pol II and its factors involved in transcription initiation, elongation and pausing, pre-mRNA processing, and transcription termination. In principle, this approach could be extended to phase complexes in which part of the structure is known and weak, anomalously scattering atoms are present.

Results and Discussion

Recovering Zn-MAD Phase Information for Pol II

Our effort to measure Zn-MAD data for phase calculation was encouraged by consistent observations of clear Zn peaks in anomalous difference Fourier maps calculated with model-derived phases (Gnatt et al., 2001; P.A.M. and J.F., unpublished data). Knowing that the ratio of zinc sites to residues of Pol II was extremely low (570 aa/Zn), which indicated only minute perturbations to the structure factors, we attempted to measure reflection amplitudes as accurately as possible with the Pol II crystals. Since the 12 subunit Pol II crystals generally diffracted weakly with the best, yet anisotropic, diffraction patterns to around 4 Å (Armache et al., 2003; Bushnell and Kornberg, 2003), and usually allowed for up to 2 complete data sets before decaying, we employed a multicrystal strategy to improve the quality of the experimental phases. In such an experiment, 5 crystals of 12 subunit Pol II were used, and only 1 (A3X7, Table 1) permitted collection of a full MAD data set with anisotropic diffraction patterns to 4.1 Å. Other crystals allowed for collection of either an anomalous dispersion data set (inflection + remote wavelength) (e.g., A3X6, Table 1) or just an anomalous peak set (A2X10, Table 1). Indicators of data quality all fell within reasonable ranges (Table 1).

The indication of feasibility to phase the large Pol II structure by using just Zn-MAD signals was first seen with dispersive difference ($F_{remote} - F_{inflection}$) Fouriers calculated with model-derived phases. In addition to strong anomalous difference ($F_{hkl} - F_{\overline{hkl}}$) peaks that corresponded to the existing Zn sites (Figure 1A), dispersive difference peaks matched precisely with the positions of zinc (Figure 1B). The dispersive peaks of eight Zn sites were observed for all of the anomalous dispersion data sets collected, except for A4X4 (where six sites were observed), indicating the presence of dispersive signals.

Next, we asked whether the crystals were isomorphous to one another at the d-spacing (resolution) to which the data sets were collected. Crosscrystal dispersive difference ($F_{remote}^{crystal-1} - F_{inflection}^{crystal-2}$) maps were calculated pairwise among the crystals. A representative map of this calculation is shown in Figure 1C. Heights of the zinc peaks decreased and noise levels increased in these maps compared to same-crystal dispersive difference maps. This was expected, as the crystals might be nonisomorphous to a certain extent. Nonetheless, the fact that zinc peaks did appear in crosscrystal dispersive difference maps encouraged the merging of phase information from individual crystals (see results below).

Experimental maps of a single phasing source of either anomalous peaks or dispersion were calculated with the known Zn atom positions. After solvent flattening, the maps showed protein boundaries that generally agreed with the known crystal structure of the 12 subunit Pol II (Figures 1D and 1E), attesting to the recording of Zn-MAD signals in the data sets. The maps based on dispersion data were weaker than those derived from the peak data (cf. Figures 1D and 1E), as expected from the general behavior of anomalous scattering. A full MAD phase set was obtained from crystal A3X7 (Table 1) up to 7.2 Å, which yielded a cleaner map that showed significantly improved agreement with the Pol II model (Figure 1F).

Further improvement of the map was explored by including MAD (SIR + SAS) phase information from other

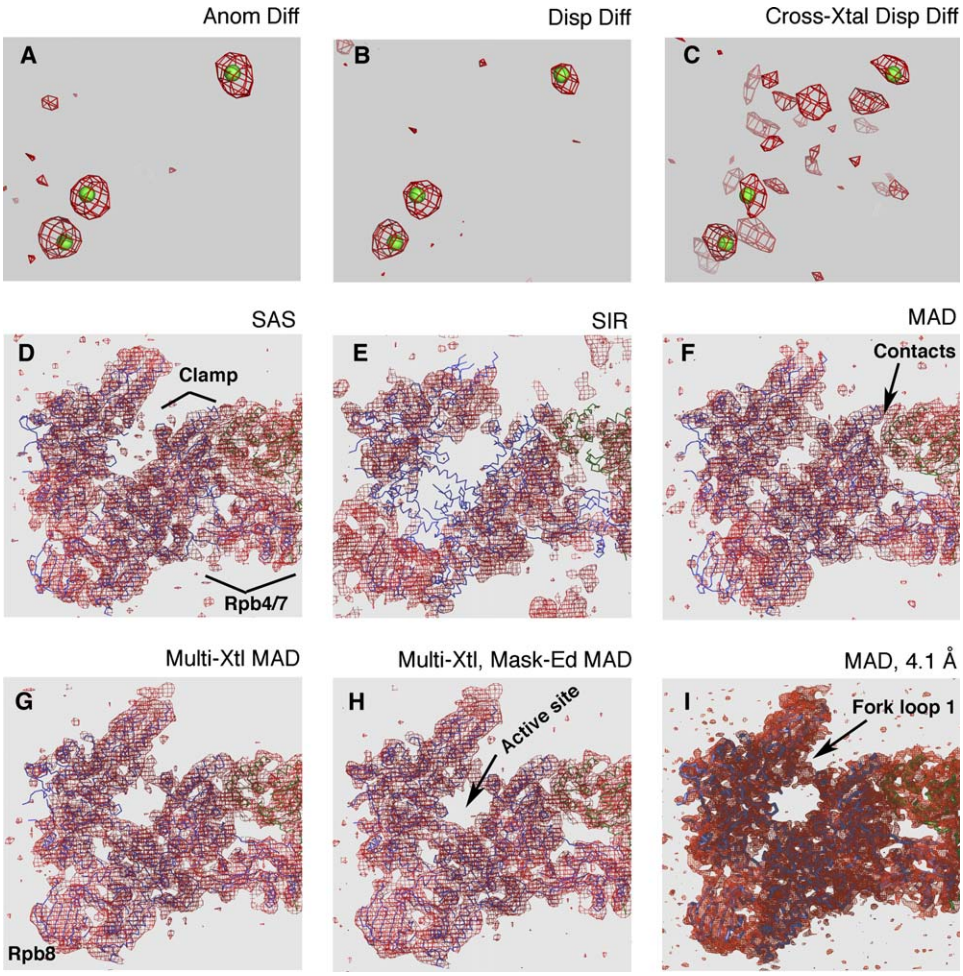

Figure 1. Calculating an Experimental Electron Density Map of Pol II by using Zn-MAD

(A) A representative anomalous difference Fourier map (7.2 Å) based on a set of Zn peak data and the model phases. Density contours are displayed at the 3.0σ (standard deviation) level in red. Three of the eight Zn sites are shown as green spheres.

(B) A representative dispersive difference Fourier map (7.2 Å) based on a set of dispersive data and the model phases. The same map parameters as in (A) were used in the calculation. The color scheme remains the same as in (A).

(C) A representative crosscrystal dispersive difference Fourier map (7.2 Å) used to assess crystal isomorphism. The map was calculated by using the model phases and amplitude differences between the remote data of crystal-1 and the inflection data of crystal-2. The map parameters and color scheme are the same as in (B).

(D–I) (D) An experimental map based solely on Zn anomalous peak data (SAS) and the known coordinates of the Zn sites. The map was calculated to 7.2 Å and is shown in red at 1.0σ. An existing 12 subunit Pol II model (Cα only) is shown in blue. The green wire model is a lattice-packing neighbor. The heterodimer of Rpb4/7 (indicated by "Rpb4/7") protrudes to the right from the bottom of the clamp domain (indicated by "Clamp"). All of the maps in (D)–(I) were solvent flattened as described in Experimental Procedures. Map slab thickness, box sizes, and the color scheme are the same in (D)–(I). (E) An experimental map calculated from a Zn dispersive data set, treated as SIR. The resolution cutoff is the same as in (D). (F) A Zn-MAD map (7.2 Å) resulted from combining the anomalous and dispersive information. Lattice contacts (indicated by "Contacts") clarify in the full MAD map, as indicated by the arrow. (G) The Zn-MAD map (7.2 Å) produced from combining multiple Zn-MAD phasing sets. The Rpb8 β sheet (indicated by "Rpb8") is labeled. (H) The Zn-MAD map resulting from the manually edited mask (for solvent flattening). The polymerase active site is indicated by the arrow. (I) The final Zn-MAD map calculated to 4.1 Å. The map was computed with smaller grid spacing than the low-resolution maps described in other panels; thus, it appears more crowded. Significantly more details of the polypeptide chains become visible in this map. The arrow indicates the density for fork loop-1 that was disordered in previous uncomplexed Pol II structures.

crystals. While two of the phasing data sets (A4X4 and A2X11) adversely influenced the final electron density map (not shown) upon phase combination, the other two (A3X6 and A2X10) improved the map quality (judged by the agreement with the model and decreased solvent noise, Figure 1G) and phasing statistics (Table 2). The MAD phase sets from the different Pol II crystals were not independent, since they were all based on the same Zn coordinates. The effect of phase improvement as observed here may be understood as follows.

In theory, upon phase combination, regions of the probability distribution that overlap will reinforce. For regions in which there is no significant overlap, the combined probability distribution will weaken. Therefore, two phase sets of moderate quality will improve upon combination of their probability distributions, whereas a poor quality set may degrade the distribution when combined with a moderate quality set. Errors in the parameters (e.g., measurement errors, nonisomorphism, or zinc atomic parameters) used to calculate the phases

Table 2. Statistics of Multicrystal Zn-MAD Phasing

Crystal Name	Phase Type	Phasing Power	F.O.M.	d-Spacing (Å)
A3X7	Dispersive	1.17	0.238	7.6
A3X7	Anomalous	2.37	0.384	7.3
A3X6	Dispersive	1.25	0.220	7.6
A2X10	Anomalous	1.93	0.345	6.5
F.O.M. after phase combination			0.560 (low resolution)	

reduce the quality of phase distribution in different ways among the data sets from different crystals; therefore, results can be improved by phase combination. The worsening of the map as a result of combining data sets from A4X4 and A2X11 crystals is attributed to significant nonisomorphism.

In the multicrystal Zn-MAD map, additional densities were observed for three regions that seemed to be disordered in the published X-ray models. For one such region of density, which was near the Rpb2 "stalk," the mask was manually expanded to allow for improved connectivity between the main body of Pol II and a large region of disconnected density (not shown). A few remaining solvent "islands" in the mask were excluded. Overall, the edited mask has a slightly larger volume (more inclusive). Solvent flattening was reperformed by using the edited mask, and it further improved the match between the map and the model (Figure 1H).

As shown in Table 1, the anomalous peak data from A3X7 was collected to 4.1 Å. The diffraction pattern beyond 6.5 Å was anisotropic and yielded an overall completeness of 85% as opposed to 99% up to 6.5 Å. Single anomalous scattering (SAS) phases were therefore available for reflections beyond the low resolution and up to 4.1 Å. However, after automatic solvent flattening at the higher resolution, the map appeared noisier (not shown) than the full Zn-MAD map at low resolution. Obviously, this was due to ambiguities associated with the SAS phases beyond the low resolution, which, in turn, did not produce the correct solvent mask. Solvent flattening was then performed by using the edited mask based on the low-resolution MAD data. Since only acentric reflections were phased in SAS, phase extension for the centric reflections was performed, again by using the low-resolution mask, to yield the final experimental electron density map of the 12 subunit Pol II at 4.1 Å (Figure 1I).

Missing Regions of the Pol II Structure Revealed by Zn-MAD

The Zn-MAD map revealed experimental densities corresponding to missing regions in the existing models of Pol II, whereas model-phased maps (F_o or $2F_o - F_c$, with or without bulk solvent correction and solvent flattening) did not show the additional densities. The first was the missing segment linking residue 437 to residue 446 that would form the tip of the Rpb2 stalk. We were able to build a loop of poly-Ala according to the density (Figure 2A). Immediately adjacent to this region, definitive electron densities (not shown) also existed that could account for two additional missing segments (residues 70–90 and 134–164 in Rpb2), but these densities were not sufficiently resolved for model building. The second location where the model was extended fell within an existing gap in Rpb4, between residues 77 and 118. Ala residues were built into the density (Figure 2B) to extend the model from residue 113 to 117, while residues 118–126 were adjusted to fit the experimental density. Partial side chain densities were clearly visible for many, but not all, residues. No model refinement was attempted since the diffraction data was limited to about 4.1 Å. Special techniques for refinement against low-resolution data (reviewed by Brunger, 2005) may be tested in the future for Pol II.

Fork loop-1 is an intriguing structural element that faces the rudder from the opposite side over the DNA/RNA cleft. It appears disordered in uncomplexed Pol II crystal structures reported to date (Cramer et al., 2001), and it adopts a stable conformation in crystals of transcription ternary complexes (Kettenberger et al., 2004; Westover et al., 2004b). Fork loop-1 interacts with one side of the rudder, and together they form two "arms" that physically hold the transcription hybrid vertical and above the catalytic center (Figure 2D). Surprisingly, we observed definitive Zn-MAD densities in the cleft of free Pol II, adjacent to the rudder. Again, model-phased maps failed to reveal the additional densities (not shown). The densities could only be accounted for by the missing fork loop-1 structure. In fact, it was not difficult to fit the densities with a loop structure (Figure 2C). The resulting conformation was compared to that in the ternary complexes of either the 10 subunit or 12 subunit Pol II. Viewed from the downstream DNA side into the catalytic center and with the cleft opening facing upward, fork loop-1 shifts toward the floor of the cleft from its position in the ternary complexes (Figure 2D). The movement at the tip of the loop is about 6 Å compared with those in the ternary complexes.

The localization of fork loop-1 indicates that the polymerase cleft is divided into two compartments, even in the absence of nucleic acids. One large void of course is the DNA/RNA hybrid tunnel that runs from the catalytic center upward (Figure 2E) for 8 base pairs (Kettenberger et al., 2004; Westover et al., 2004b). The other half of the cleft lies nearly perpendicular down the hybrid tunnel and forms an extensive channel for binding the downstream DNA duplex (Figure 2E, also Figure 4 of Gnatt, 2002). This compartmentalization of the unoccupied cleft argues that fork loop-1 as well as the rudder must change conformation transiently upon interaction with template DNA, and that, consequently, they can open the hybrid tunnel. A swinging motion of the clamp domain is most likely involved in the event. The template strand is then embedded in the tunnel by the resumption of stable conformations of the loops. After initiation, the loops interact with RNA in the hybrid (Westover et al., 2004b), and that may further stabilize the conformation of fork loop-1. Also important is the network of interactions from fork loop-1 to the rudder and from the rudder to the lid (the lid serves to separate the RNA and DNA strands), as has been discussed (Boeger et al., 2005; Westover et al., 2004b). Interactions between the tip of fork loop-1 (Rpb2, Lys471–Met473) and one side of the rudder (Rpb1, Leu315, Ser318–Pro321) (Figure 2C) are crucial to the retention of nucleic acids in the hybrid tunnel, as deletion of two loop-1 residues (Rpb2, 471–472) resulted in loss of initiation activity (Jeronimo et al.,

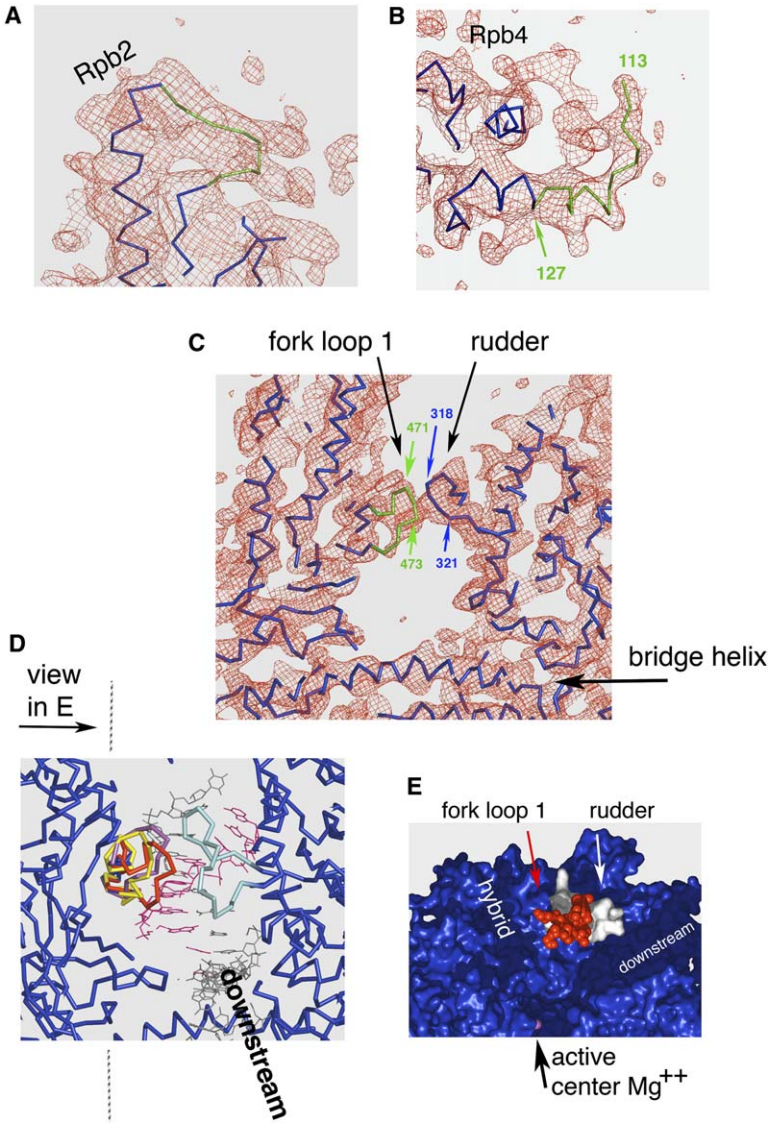

Figure 2. Updating the Pol II Model According to the Zn-MAD Densities

(A–C) (A) Definition of the loop structure at the tip of a two-helix stalk in Rpb2. Poly-alanines (green, Cα only) were fitted into the density, connecting the helices. Electron densities (red) are shown all at 1.0σ level for (A)–(C). The existing model of Pol II (Cα only) is colored blue in (A)–(C). Some of the side chain densities that projected off the α carbons were evident in the 4.1 Å Zn-MAD map. (B) The model for Rpb4 was updated according to the Zn-MAD densities. Residues 118–126 were adjusted to fit the density. Residues 113–117 were inserted according to the density. α carbons of the updated residues are shown in green. Again, partial densities of certain side groups protrude from the α carbons. (C) Localization of fork loop-1 (green, Cα model) according to its Zn-MAD density. Clear densities occurred at a site opposite the rudder over the cleft, and they could only be accounted for by fitting with fork loop-1. As shown by the touching densities, fork loop-1 contacts the rudder to form a set of "arms" that surrounds the DNA/RNA hybrid in the ternary complex as has been noted (Westover et al., 2004b). Residues involved in the contact are indicated. The bridge helix defining the floor of the cleft is indicated as well.

(D and E) (D) A motion by fork loop-1 can be seen from its different locations in the free Pol II (red, Cα model) and in ternary complexes (yellow and purple). The view is roughly from the downstream side through the cleft and against the wall (not visible here) of the hybrid tunnel. The rudder is shown in cyan, and the rest of Pol II is shown in blue. The DNA template strand is shown by the gray stick model, while the transcript RNA is indicated by the pink sticks. The downstream DNA (gray sticks) is roughly perpendicular to the plane of the paper. The dashed line indicates a cut-away plane against which is the viewing direction for (E). (E) Compartmentalization of the Pol II cleft by the protein mass of fork loop-1 (red) and the rudder (white). This is a cut-away view from the plane, as indicated in (D), against the clamp domain of Rpb1. The surface rendering shows the spatial arrangement of the hybrid tunnel and the downstream DNA channel.

2004). Although the deletion mutant was able to enter a preinitiation complex in the presence of general transcription factors (Jeronimo et al., 2004), it was not able to stabilize the open promoter complex, presumably because it could not help embed the template DNA and the hybrid in the manner prescribed by the interaction network.

The strand-loop network of interactions may be considered to be an intramolecular "relay" system that is also sensitive to structural perturbations induced by external factors such as the 3′ processing/termination factor Pcf11 and the monoclonal antibody 8WG16, which specifically recognizes the C-terminal repeated domain (CTD) of Rpb1. Even though these factors do not directly access the DNA/RNA cleft, incubation with them resulted in disassembly of paused ternary complexes (Zhang et al., 2004, 2005), suggesting a structural mechanism that propagates actions from the external factors to the center of the polymerase cleft. The linker sequence in the Pol II structure that leads to the CTD may be a component of the mechanism, since the CTD mediates the interaction with these factors (Meinhart and Cramer, 2004; Thompson et al., 1990). Alternatively, binding of external factors to the N-terminal part of CTD, which is proximal to the clamp, may trigger outward movement of the clamp, which, in turn, releases the interaction of its rudder domain with fork loop-1, leading to the opening of the DNA/RNA tunnel.

Effectiveness of Multicrystal Zn-MAD Phasing

MAD phases based on anomalous scattering from Zn atoms have been utilized to solve protein and nucleic acid structures (summarized in Table 3). To date, the most effective phasing by intrinsically bound Zn atoms was found in the crystallographic work on SV40 large T-antigen (Li et al., 2003), where the 368 residue

Table 3. Macromolecular Structures Recently Solved with Zn-MAD Phasing

Molecule	MW[a]; No. Res. (ASU[b])	No. Zn (ASU)	Phasing Effectiveness	Phasing Method; Resolution	Reference
Ketose Reductase	38,151 Da; 352 aa	2	151 aa/Zn	MAD; 2.3 Å	Banfield et al., 2001
HIV-1 DIS	15,722 Da; 46 nt	5	9 nt/Zn	MAD; 2.0 Å	Ennifar et al., 2001
SV40 Large T-antigen	84,866 Da; 736 aa	2	368 aa/Zn	MAD; 2.8 Å	Li et al., 2003
Tetanus Neurotoxin Light Chain	51,158 Da; 447 aa	2	212 aa/Zn	MAD/MR[c]; 2.3 Å	Breidenbach and Brunger, 2005
Acyl Carrier Protein	8,500 Da; 77 aa	3	26 aa/Zn	MAD; 1.1 Å	Qiu and Janson, 2004
RNA Polymerase II	513 kDa; 4,565 aa	8	570 aa/Zn	MAD; 4.1 Å	Present study

[a] MW, molecular weight.
[b] ASU, crystallographic asymmetric unit.
[c] MR, molecular replacement.

structure was determined by using a single Zn site. Our results show that this is not the limit on the mass of protein that can be phased crystallographically by one Zn site. Rather, a phasing effectiveness of 570 amino acids per Zn atom is achieved here (Table 3).

As described above, it is essential that redundant phasing sets be collected for recovering the relatively weak anomalous scattering effects from zinc. In addition, the Zn-MAD approach requires prior knowledge of the structural model, because the Zn peaks were only revealed in difference Fourier maps (Figures 1A and 1B) with model phases; they were "invisible" in anomalous difference Patterson maps. To investigate the degree of prior information required for this approach, varying parts of the Pol II structure were omitted during phase calculations for anomalous difference Fourier maps. Based on the resulting difference peaks, we were able to conclude that models as partial as 50% were adequate to reveal the Zn positions (not shown).

With the currently available coordinates of Pol II (Armache et al., 2005) and its complexes (Kettenberger et al., 2003, 2004), new structures of regulated complexes of Pol II may be solved conveniently in similar Zn-MAD experiments. In principle, the same approach can be applied to solving any large Zn (or other weak anomalous scatterer)-containing complex whose partial crystallographic model is available. Based on our experience with Pol II and computational experiments described in the next section, we recommend the following procedure/guidelines for the treatment of MAD data:

(1) As a general limit of phasing effectiveness, 1 Zn site may phase up to 1100 residues of amino acids;
(2) Acquire MAD data sets (i.e., peak, inflection, and remote) accurately by using high-redundancy collection strategies;
(3) Use multiple crystals in order to gain redundancy in terms of individual phase sets;
(4) Generate a master amplitude file by merging all of the individual diffraction data sets from various crystals; Scaling individual phasing sets to the master set is fundamentally important to the calculation of multicrystal MAD phases and can be carried out as described below;
(5) Scale all data sets at the *peak* wavelength to the master amplitudes; scale *inflection* data sets to the master as well; and, importantly, scale individual *remote* sets to the prescaled inflection set from the same crystal;
(6) Calculate Zn-MAD phases for individual crystals separately by treating the data as SIRAS (Ramakrishnan and Riou, 1997);
(7) Finally, calculate phases by combining Zn-MAD phase sets from individual crystals. Selection of an optimal combination of phase sets from various crystals is critical to the success of phasing by this approach and must be carried out systematically by using the resulting electron density map (solvent-flattened) as a judgement.

Estimating the Limit of Multicrystal Zn-MAD Phasing

Now, we address the issue concerning the upper limit of phasing effectiveness by Zn-MAD. In other words, how many amino acids of polypeptide can be phased by a single Zn site in multicrystal MAD experiments? Due to the absence of a crystallizable protein having a Zn content more "dilute" than that of Pol II, we performed computational experiments by using simulated diffraction amplitudes calculated from Zn-containing (varied from eight sites to one site) Pol II models with added random errors (see Experimental Procedures for details).

For a quantitative control on approximation by the simulation, we first tested the Pol II model with a full Zn content (eight sites). We found that Zn-MAD data simulated for two crystals yielded maps comparable to the experimental ones, with regard to the heights of Zn peaks in anomalous difference Fourier syntheses (Table S2; see the Supplemental Data available with this article online), the noise level in the resulting protein map, and coverage of the Pol II molecule (Figure S1, cf. [A] and [B]). Thus, the procedures used for error generation had simulated experimental conditions of the weakly diffracting Pol II crystals. Because the random errors were added directly to the structure factor amplitudes, a sum of effects had been simulated, including counting statistics, nonisomorphism, absorption, wavelength impurity, and data scaling and reduction errors.

In terms of the map correlation coefficient (C.c) and the real space R factor (RSR) (Table S1), our simulation shows that as the number of Zn decrements from eight to two sites per Pol II molecule, more individual crystals (or individual MAD data sets) are required in order to produce electron density maps that match the model. For instance, with four Zn atoms per Pol II, five crystals will be needed in order to produce a map with the same quality as that for eight Zn sites and two crystals (Figure S1C), whereas two crystals of Pol II with four Zn sites prove insufficient (Figure S1D). No reliable map could be

produced when the Zn content dropped to one per Pol II molecule (Table S1; maps not shown).

We note in Table S1 that the simulated MAD experiments deviate somewhat from a linear relationship between the number of Zn sites and the minimum number of crystals required. We think this fluctuation of phasing behavior is due to the inclusion of certain "nonisomorphous" crystal(s) in the simulated MAD data sets. For optimal MAD phase calculations, data from such crystals should be excluded in real-case experiments (see guideline 7 above). We can also see from the computational experiments that a minimum "concentration" of 2–3 sites per Pol II molecule (corresponding to over 1600 aa/Zn) is required for an experiment with 10 crystals or more. However, if we consider the collection of five individual MAD data sets to be the practical limit that a single experiment may achieve (considering factors such as synchrotron beam-time allocation and crystal availability), it will then require at least four Zn sites per Pol II. Accordingly, we conclude that 1 Zn site is able to provide experimental phase information for as many as 1100 amino acids of polypeptide, under the condition of the minimal errors achievable with current synchrotron and detector technologies.

Alternative Multicrystal Approaches

In principle, the multicrystal Zn-MAD approach described above should serve well for phasing large proteins by using the Se-methionine substitution strategy, since selenium and zinc have anomalous scattering factors with comparable magnitudes ($f'' \approx 4.5$). Coordinated by special polypeptide motifs, intrinsically bound Zn ions are generally well localized and thus will give rise to strong anomalous difference peaks and effective phasing power, whereas certain Se sites may be disordered depending on the mobility or conformational homogeneity of the side groups of substituted methionines. Nonetheless, we believe that multiwavelength anomalous scattering from well-ordered Se-methionines should prove as effective as MAD from Zn. Application of this approach to experiments based on other weak anomalous scatterers such as sulfur and calcium (Ramagopal et al., 2003; Roeser et al., 2005), however, is likely to encounter problems of very small f'' values and significant crystal absorption, as dictated by X-ray absorption spectra of the elements.

Finally, we investigated the analogous approach of multicrystal SAS phasing, using our data collected at the Zn absorption peak. We found that, although this produced slight improvements compared to single-crystal SAS, the resulting maps were clearly of lower quality than multicrystal MAD maps (not shown), most likely due to unresolved phase ambiguity.

Experimental Procedures

Protein Preparation and Crystallization

Yeast RNA polymerase II was purified directly from a haploid strain of *S. cerevisiae* by using the TAP-tagging strategy (Rigaut et al., 1999) and procedures essentially as described previously (Bushnell and Kornberg, 2003). Purified polymerase samples were concentrated to 3.5 mg/ml in 5 mM Tris-HCl (pH 7.5), 40 mM $(NH_4)_2SO_4$, 10 mM DTT, and a cocktail of protease inhibitors. Crystals were grown in hanging drops composed of 2 μl of the Pol II sample and 2 μl of the mother liquor (5%–7% PEG4000, 100 mM MgAc, 50 mM HEPES [pH 7.0], 10 mM DTT, 70 mM glucosamine, and 50 mM Dioxane). Crystals (0.1–0.2 mm) were adapted to a cryoprotection solution composed of the mother liquor with a higher concentration of PEG4000 (10%) plus 30%–36% glycerol. Cryosoaked crystals were flash frozen and stored in liquid nitrogen; they were kept under a cold nitrogen (100 K) gas stream during diffraction experiments.

Data Collection and Reduction

Several factors influenced the data collection strategy. Most of the 12 subunit Pol II crystals diffracted only to low resolutions (e.g., 7 Å) and were radiation sensitive. X-ray fluorescence scans clearly showed the presence of zinc in the crystals. Spectra from the scans (not shown) were nearly identical to published data (Ennifar et al., 2001), except the estimated anomalous scattering factors were slightly smaller ($f' = -9.21 \pm 0.69$; $f'' = 4.48 \pm 0.54$, for 26 scans of 6 crystals). In most cases, it was not possible to collect a full MAD data set from a single crystal. In the only case where it was possible, images collected later in the process showed radiation damage evidenced by increased mosaicity and decreased resolution. In general, individual crystals were used for either anomalous or dispersive data, but not for both. Due to the relative weakness of the Zn signals and to radiation sensitivity of the crystals, data sets were collected in wedges of 20°. For dispersive data sets, two wedges consisting of the same oscillation range were used sequentially for the inflection and remote wavelength. For anomalous data, wedges were collected by using the standard "inverse-beam" arrangement (collect a wedge starting at n°, followed by a wedge starting at n° + 180°). One dispersive data set (A4X4, Table 1) was not collected in wedges and was later found to yield poor-quality phases. The weakness of the zinc signals also dictated that as many measurements as possible should be taken for any given reflection. Multiple data sets were thus collected, and each data set was measured to full completeness. The highest resolution observed was anisotropic diffraction to 4.1 Å from crystal A3X7 (Table 1), which was collected at the wavelength for the Zn anomalous peak. This data set was treated up to 6.5 Å first, and the higher-resolution information was incorporated later. From five crystals, a total of 11 data sets, with inflection and remote sets counted separately, were collected (Table 1).

All indexing and integration steps were performed by using the stand-alone version (6.2.3) of MOSFLM (Leslie, 1992) modified to allow output of the orientation matrix. Unit cell parameters derived from different crystals, data sets, or sets of images within a data set differed by as much as 5 Å, complicating the determination of isomorphism. Indexing was usually achieved by using three images at n°, n° + 45°, and n° + 90° and refined by using three batches of images at each of those positions. Integration was performed in the same wedges as data collection. For anomalous data sets, the wedges were integrated in chronological order rather than rotation order. In order to reduce the effect of small changes in the orientation matrix caused by the inversion of oscillation ranges, the current orientation matrix was saved after completing integration of a wedge. The orientation matrix was reloaded and used for integration of the subsequent (rotation order) wedge. This improved the agreement between predicted and observed spot positions. Later, comparisons of Zn-SAS maps and anomalous difference Fouriers showed that integration in chronological order was superior to integration in rotation order. Chronological order integration could be thought of as allowing smooth variation of the parameters within MOSFLM to model the change in the X-ray beam over time, producing fewer abrupt changes that could cause problems for refinement.

Scaling of data sets was done with SCALA, part of the CCP4 suite (version 4.2.2) (CCP4, 1994); each wedge was treated as a separate run. Additionally, all diffraction data sets were scaled together into two "master" data sets, one up to 6.5 Å and the other to 4.1 Å (Table 1). The master data sets could not be used for phasing, as the amplitude differences needed for phase calculation had been averaged out. However, this was technically important, as the master sets provided reference for scaling the individual data sets prior to phase calculation and combination.

Test of Isomorphism at Low Resolution

As mentioned above, the unit cell parameters based on the low-resolution data did not permit reliable determination of isomorphism between crystals. Isomorphism was then tested with dispersive

difference Fourier by using remote wavelength data from one crystal and inflection data from another crystal. Although there was more noise present in crosscrystal dispersive difference maps than in single-crystal dispersive maps, the same Zn positions were observed consistently for pairs of the crystals (Figure 2C).

Locating Zinc Sites and Initial Estimation of Phase Quality

The eight zinc sites of Pol II were located by anomalous or dispersive difference Fourier maps. The phases used for these maps were calculated from an existing 12 subunit Pol II model (PDB code: 1WCM) after rigid-body refinement against experimental amplitudes with CNS (Brunger et al., 1998). The known mobile domains of Pol II (Cramer et al., 2001) were used as rigid bodies. Four additional domains were further defined: each of the Pol II "feet" (Rpb8 and Rpb1 877–1057) was treated as a separate rigid domain, as was Rpb5; the Rpb4 and Rpb7 heterodimers were treated as a single rigid body.

Each phasing data set was initially treated as SAS ($F_{hkl} - F_{\overline{hkl}}$) or SIR ($F_{remote} - F_{inflection}$) (Ramakrishnan and Riou, 1997), and solvent flattening was performed with the PHASES package (Furey and Swaminathan, 1997). The quality of each phasing data set could be judged by the levels of noise in the difference and SAS/SIR F_o maps and the peak heights in the difference maps. Crossdifference maps calculated by using phases from one phasing data set (e.g., SAS phases instead of model phases) with amplitude differences from another source showed peaks in the same positions, which corresponded with the known positions of the Zn sites. The difference maps also indicated that the SAS phases were generally of higher quality than the dispersive phases (treated as SIR), which was confirmed by comparison of the SAS and SIR F_o maps after solvent flattening.

Prior to calculation of SIR/SAS maps, the Zn sites were refined against the anomalous or dispersive differences accordingly, by using GREF of PHASES. During phase calculation, maximum likelihood refinement against a lack-of-closure error was invoked. Possibly due to the resolution of the data, lack-of-closure refinement of occupancy and temperature factor was not stable.

Phase Calculation and Improvement

Zn-MAD phase values were all calculated with PHASES by using the option of maximum likelihood refinement. Dispersive differences were treated as SIR. After a combinatorial test, two sets of SAS phases (Zn peak) and two sets of dispersive phases from a total of three crystals were combined in the final calculation, while two other dispersive sets (A4X4 and A2X11, Table 1) were discarded since they worsened the map. The statistics of the procedure are given in Table 2. Another phase set was produced similarly but with the inclusion of anisotropic SAS information to 4.1 Å from crystal A3X7 (Table 1). Maximum likelihood refinement of the 4.1 Å phases was unstable, so the refined zinc parameters from the low-resolution phase set were used. Although model phases were available, they were not used as an external reference phase set during phase refinement, so as to exclude any possible influence from the model.

The resulting multicrystal Zn-MAD phase set to low resolution was subjected to automatic solvent flattening for phase improvement by using the standard parameter values with the solvent density value set to 0.25 with a smearing radius of 14.0 Å. The resulting electron density map and the associated protein envelope were examined for agreement with the known structure of the 12 subunit Pol II. The solvent mask was converted to the compressed "O" mask format and edited manually in "O" to allow for connectivity in one of the regions (Rpb2 residues 70–90 and 134–164) where there were breaks in the Pol II model. No remodeling was performed in this region, however. The mask was also edited to exclude a few islands of solvent noise. No mask modification was applied to the newly modeled regions (three in total). The strength of experimental phases was tested by setting the mask region containing Rpb5 to "solvent," and this region of density was observed to return after solvent flattening, although it was slightly weakened. Upon verification against the known model, the polymerase envelope derived at low resolution was used in another round of solvent flattening on the phase set that also included the SAS phases from crystal A3X7, thus producing the final experimental electron density map to 4.1 Å. We verified that changing the lower-limit cutoff (120 Å, 50 Å, and 25 Å) in map calculation had virtually no effect on the final map. Finally, the averaged phase angle difference between the model and the solvent-flattened MAD was 68° at low resolution and 73° at 4.1 Å. These values seem slightly greater than in ideal cases, but they are attributable to the fact that the 12 subunit Pol II model has a relatively low percentage (67% versus 90% ideally) of residues falling within the most favored peptide dihedral angle regions.

Model Building

Model building for three regions of the structure was performed with the molecular graphics program "O" (Jones et al., 1991). Models for side chains were not attempted, although densities were present for the side groups of some residues at 4.1 Å. Due to the limited resolution, no attempts were made to refine the updated model. Structural images were generated with PyMOL (DeLano, 2002).

Simulation of MAD Structure Factor Amplitudes

Error-free structure factor amplitudes were initially calculated to 4 Å for individual wavelengths that corresponded to the peak, inflection, and remote wavelengths of the zinc absorption spectrum, with the inclusion of anomalous scattering factors for elements in the protein (Zn, C, O, N, S). More precisely, normal structure factors were modified by the anomalous contributions ($\Delta f' + i\Delta f''$) corresponding to the wavelengths.

For the simulation to be as close to experimental conditions as possible, the error-free structure factor amplitudes (F^{sim}) were scaled to the experimental amplitudes of the Pol II master data set, and sigma values ($\sigma[F^{sim}]$) for the calculated amplitudes were then taken from the master data set ($\sigma[F^{obs}]$). To generate F^{sim} with random errors, F^{sim} and $\sigma(F^{sim})$ were converted to I^{sim} and $\sigma(I^{sim})$. Random errors to I^{sim} were generated by using Gaussian distributions with $\sigma(I^{sim})$ as standard deviations (Box and Muller, 1958). Alternatively, we used the empirical approach for generating sigma values as a function of the magnitude of reflection intensities (Hubbard et al., 1995). However, the approach did not produce consistent results further along the test, most likely due to the poor fit of the quadratic equation to experimental data in the resolution range available; therefore, it was not pursued. The randomly modified I^{sim} values were then converted back to produce error-containing F^{sim}. To simulate diffraction from multiple crystals, the same procedure was repeated so that multiple sets of amplitudes were generated, each having varying modifications by the random value generator described above. Since all of the F^{sim} values were based on the model, no further scaling was necessary.

We first examined the simulated anomalous difference Fourier for eight Zn sites, corresponding to the Zn content in Pol II. We found that it was necessary to additionally scale down the errors by a factor of 4.0 in order to produce simulated Zn peaks comparable in height to the observed ones (Table S2). As such, the simulated data sets from two crystals, when subjected to the same phasing procedure as used for the experimental data, resulted in a protein map comparable to the experimental Zn-MAD map (Figures S1A and S1B). Therefore, the same reduction was applied throughout the data simulation for the various Zn contents.

Map Comparison

Map comparisons were first done by visual inspections. For quantitative comparisons between the simulated maps and the Pol II model, correlation coefficients and real space R factors were calculated (Table S1). The correlation coefficients were computed as:

$$C.c. = \frac{<\rho_{sim}(r)\cdot\rho_{cal}(r)> - <\rho_{sim}(r)>\cdot<\rho_{cal}(r)>}{\sqrt{<\rho^2_{sim}(r)> - <\rho_{sim}(r)>^2}\cdot\sqrt{<\rho^2_{cal}(r)> - <\rho_{cal}(r)>^2}}, \quad (1)$$

with r being a coordinate (grid point) in the unit cell, and $\rho_{sim}(r)$ and $\rho_{cal}(r)$ being densities corresponding to simulated maps and the map calculated from the model, respectively (Lunin and Woolfson, 1993). Real space R factors were calculated by using the formula:

$$RSR = \frac{\sum_r |\rho_{sim}(r) - \rho_{cal}(r)|}{\sum_r |\rho_{sim}(r) + \rho_{cal}(r)|}, \quad (2)$$

where the summations were carried over the entire range of grid points (r) of the asymmetric unit (Drenth, 1999).

Supplemental Data

Supplemental data include Figure S1, and Tables S1 and S2, and can be found with this paper online at http://www.structure.org/cgi/content/full/14/6/973/DC1/.

Acknowledgments

This work was supported by research grants from the National Institutes of Health (GM064651) and the American Cancer Society (IRG-01-231-01) to J.F. We thank Shai Vaday and Michael Becker of National Synchrotron Light Source (NSLS) for technical assistance. We are grateful to David Schuller of the Macromolecular Diffraction Facility at the Cornell High Energy Synchrotron Source (MacCHESS) for maintaining computer systems. Use of the NSLS, Brookhaven National Laboratory, was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-98CH10886. This work is also based on research conducted at CHESS, which is supported by the National Science Foundation and the National Institutes of Health/National Institute of General Medical Sciences under award DMR-0225180.

Received: October 5, 2005
Revised: February 23, 2006
Accepted: April 19, 2006
Published: June 13, 2006

References

Armache, K.J., Kettenberger, H., and Cramer, P. (2003). Architecture of initiation-competent 12-subunit RNA polymerase II. Proc. Natl. Acad. Sci. USA *100*, 6964–6968.

Armache, K.J., Mitterweger, S., Meinhart, A., and Cramer, P. (2005). Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. J. Biol. Chem. *280*, 7131–7134.

Ban, N., Freeborn, B., Nissen, P., Penczek, P., Grassucci, R.A., Sweet, R., Frank, J., Moore, P.B., and Steitz, T.A. (1998). A 9 Å resolution X-ray crystallographic map of the large ribosomal subunit. Cell *93*, 1105–1115.

Ban, N., Nissen, P., Hansen, J., Capel, M., Moore, P.B., and Steitz, T.A. (1999). Placement of protein and RNA structures into a 5 Å-resolution map of the 50S ribosomal subunit. Nature *400*, 841–847.

Banfield, M.J., Salvucci, M.E., Baker, E.N., and Smith, C.A. (2001). Crystal structure of the NADP(H)-dependent ketose reductase from *Bemisia argentifolii* at 2.3 Å resolution. J. Mol. Biol. *306*, 239–250.

Boeger, H., Bushnell, D.A., Davis, R., Griesenbeck, J., Lorch, Y., Strattan, J.S., Westover, K.D., and Kornberg, R.D. (2005). Structural basis of eukaryotic gene transcription. FEBS Lett. *579*, 899–903.

Box, G.E.P., and Muller, M.E. (1958). A note on the generation of random normal deviates. Ann. Math. Stat. *29*, 610–611.

Breidenbach, M.A., and Brunger, A.T. (2005). 2.3 Å crystal structure of tetanus neurotoxin light chain. Biochemistry *44*, 7450–7457.

Brunger, A.T. (2005). Low-resolution crystallography is coming of age. Structure *13*, 171–172.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

Bushnell, D.A., and Kornberg, R.D. (2003). Complete, 12-subunit RNA polymerase II at 4.1-Å resolution: implications for the initiation of transcription. Proc. Natl. Acad. Sci. USA *100*, 6969–6973.

Bushnell, D.A., Cramer, P., and Kornberg, R.D. (2002). Structural basis of transcription: α-amanitin-RNA polymerase II cocrystal at 2.8 Å resolution. Proc. Natl. Acad. Sci. USA *99*, 1218–1222.

Bushnell, D.A., Westover, K.D., Davis, R.E., and Kornberg, R.D. (2004). Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Å. Science *303*, 983–988.

CCP4 (Collaborative Computational Project, Number 4) (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. *50*, 760–763.

Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., and Kornberg, R.D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. Science *288*, 640–649.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. Science *292*, 1863–1876.

Crick, F., and Maydoff, B. (1956). The theory of the method of isomorphous replacement for protein crystals. I. Acta Crystallogr. *9*, 901–908.

Darst, S.A., Edwards, A.M., Kubalek, E.W., and Kornberg, R.D. (1991). Three-dimensional structure of yeast RNA polymerase II at 16 Å resolution. Cell *66*, 121–128.

DeLano, W.L. (2002). The PyMOL Molecular Graphics System (San Carlos, CA: DeLano Scientific).

Drenth, J. (1999). Principles of Protein X-Ray Crystallography, Second Edition (New York: Springer-Verlag).

Edwards, A.M., Darst, S.A., Feaver, W.J., Thompson, N.E., Burgess, R.R., and Kornberg, R.D. (1990). Purification and lipid-layer crystallization of yeast RNA polymerase II. Proc. Natl. Acad. Sci. USA *87*, 2122–2126.

Edwards, A.M., Kane, C.M., Young, R.A., and Kornberg, R.D. (1991). Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. J. Biol. Chem. *266*, 71–75.

Ennifar, E., Walter, P., and Dumas, P. (2001). An efficient method for solving RNA structures: MAD phasing by replacing magnesium with zinc. Acta Crystallogr. D Biol. Crystallogr. *57*, 330–332.

Fu, J., Gnatt, A.L., Bushnell, D.A., Jensen, G.J., Thompson, N.E., Burgess, R.R., David, P.R., and Kornberg, R.D. (1999). Yeast RNA polymerase II at 5 Å resolution. Cell *98*, 799–810.

Furey, W., and Swaminathan, S. (1997). PHASES-95: a program package for processing and analyzing diffraction data from macro-molecules. Methods Enzymol. *277*, 590–620.

Gnatt, A. (2002). Elongation by RNA polymerase II: structure-function relationship. Biochim. Biophys. Acta *1577*, 175–190.

Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. Science *292*, 1876–1882.

Hartzog, G.A. (2003). Transcription elongation by RNA polymerase II. Curr. Opin. Genet. Dev. *13*, 119–126.

Hendrickson, W.A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. Science *254*, 51–58.

Hendrickson, W.A., and Ogata, C.M. (1997). Phase determination from multiwavelength anomalous diffraction measurements. Methods Enzymol. *276*, 494.

Hubbard, S.R., Greenall, R.J., and Woolfson, M.M. (1995). On the application of anomalous scattering in oligonucleotide crystallography. Acta Crystallogr. D Biol. Crystallogr. *51*, 979–989.

Jeronimo, C., Langelier, M.F., Zeghouf, M., Cojocaru, M., Bergeron, D., Baali, D., Forget, D., Mnaimneh, S., Davierwala, A.P., Pootoolal, J., et al. (2004). RPAP1, a novel human RNA polymerase II-associated protein affinity purified with recombinant wild-type and mutated polymerase subunits. Mol. Cell. Biol. *24*, 7043–7058.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.

Kettenberger, H., Armache, K.J., and Cramer, P. (2003). Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. Cell *114*, 347–357.

Kettenberger, H., Armache, K.J., and Cramer, P. (2004). Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. Mol. Cell *16*, 955–965.

Kornberg, R.D. (1999). Eukaryotic transcriptional control. Trends Cell Biol. *9*, M46–M49.

Lattman, E. (1985). Use of the rotation and translation functions. Methods Enzymol. *115*, 55–77.

Leslie, A.G.W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. CCP4 Newsletter on Protein Crystallography, No. 26.

Li, D., Zhao, R., Lilyestrom, W., Gai, D., Zhang, R., DeCaprio, J.A., Fanning, E., Jochimiak, A., Szakonyi, G., and Chen, X.S. (2003). Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen. Nature *423*, 512–518.

Lunin, V.Y., and Woolfson, M.M. (1993). Mean phase error and the map-correlation coefficient. Acta Crystallogr. D Biol. Crystallogr. *49*, 530–533.

Meinhart, A., and Cramer, P. (2004). Recognition of RNA polymerase II carboxy-terminal domain by 3′-RNA-processing factors. Nature *430*, 223–226.

Qiu, X., and Janson, C.A. (2004). Structure of apo acyl carrier protein and a proposal to engineer protein crystallization through metal ions. Acta Crystallogr. D Biol. Crytallogr. *60*, 1545–1554.

Ramagopal, U.A., Dauter, M., and Dauter, Z. (2003). Phasing on anomalous signal of sulfurs: what is the limit? Acta Crystallogr. D Biol. Crystallogr. *59*, 1020–1027.

Ramakrishnan, V., and Riou, V. (1997). Treatment of multiwavelength anomalous diffraction data as a special case of multiple isomorphous replacement. Methods Enzymol. *276*, 538–557.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. Nat. Biotechnol. *17*, 1030–1032.

Roeder, R.G. (1996). The role of general initiation factors in transcription by RNA polymerase II. Trends Biochem. Sci. *21*, 327–335.

Roeser, D., Dickmanns, A., Gasow, K., and Rudolph, M.G. (2005). De novo calcium/sulfur SAD phasing of the human formylglycine-generating enzyme using in-house data. Acta Crystallogr. D Biol. Crystallogr. *61*, 1057–1066.

Rossmann, M.G. (2001). Molecular replacement—historical background. Acta Crystallogr. D Biol. Crystallogr. *57*, 1360–1366.

Shilatifard, A. (2004). Transcriptional elongation control by RNA polymerase II: a new frontier. Biochim. Biophys. Acta *1677*, 79–86.

Thompson, N.E., Aronson, D.B., and Burgess, R.R. (1990). Purification of eukaryotic RNA polymerase II by immunoaffinity chromatography. Elution of active enzyme with protein stabilizing agents from a polyol-responsive monoclonal antibody. J. Biol. Chem. *265*, 7069–7077.

Westover, K.D., Bushnell, D.A., and Kornberg, R.D. (2004a). Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. Cell *119*, 481–489.

Westover, K.D., Bushnell, D.A., and Kornberg, R.D. (2004b). Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. Science *303*, 1014–1016.

Woychik, N.A., and Young, R.A. (1990). RNA polymerase II: subunit structure and function. Trends Biochem. Sci. *15*, 347–351.

Zhang, Z., Wu, C.H., and Gilmour, D.S. (2004). Analysis of polymerase II elongation complexes by native gel electrophoresis: evidence for a novel carboxyl-terminal domain-mediated termination mechanism. J. Biol. Chem. *279*, 23223–23228.

Zhang, Z., Fu, J., and Gilmour, D.S. (2005). CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3′-end processing factor, Pcf11. Genes Dev. *19*, 1572–1580.

Accession Numbers

The updated coordinates of yeast Pol II, along with full structure factors (amplitudes and Zn-MAD phases), have been deposited in the Protein Data Bank with accession code 2B8K.